

Construct validity in Operations Management by using Rasch Measurement Theory. The case of the construct “motivation to implement continuous improvement”

Validación de constructos en dirección de operaciones mediante la Teoría de la Medición de Rasch. El caso del constructo “motivación para implantar la mejora continua”

Lidia Sanchez-Ruiz^a y Beatriz Blanco^b

^aDpto. Administración de Empresas, Universidad de Cantabria, Spain, sanchezrl@unican.es; ^bDpto. Administración de Empresas, Universidad de Cantabria, Spain, blancob@unican.es

Recibido: 2016-07-21 Aceptado: 2016-09-12

Abstract

Construct design and validation is a common practise in the Operations Management field. In this sense, the aim of this study is to present Rasch Measurement Theory (RMT) as a richful and useful methodology in order to validate constructs. In order to do so, the measurement controversy in the social science is presented; then, RMT is explained as a solution for this measurement issue; after that, the different applications of RMT are described and, finally, the different stages of the validation process are presented. Thus, this work aims to serve as a guide for those researchers interested in the methodology. Therefore, a specific case is included: the validation of the construct “motivation to implement continuous improvement”.

Keywords: Rasch Measurement Theory; construct; validation; reliability; item; subject

Resumen

El diseño y validación de constructos es una práctica habitual en el campo de investigación de la Dirección de Operaciones. En esta línea, el objetivo de este trabajo es presentar la Teoría de la Medición de Rasch (TMR) como metodología para la validación de constructos. Para ello se describe la problemática de medición concreta de las ciencias sociales, cómo la Teoría de la Medición de Rasch da respuesta a esa problemática, las distintas aplicaciones de la TMR, así como las diferentes etapas a realizar para comprobar si un constructo es válido o no. Así, el trabajo pretende servir de guía para todos aquellos interesados en la metodología. Para ello se ofrece un caso concreto de aplicación al constructo “motivación para implantar la mejora continua”.

Palabras clave: Teoría de la Medición de Rasch; constructo; validación; fiabilidad; item; sujeto

1. Introducción

El diseño y validación de constructos es una práctica necesaria y habitual en el campo de investigación de la Dirección de Operaciones. Son muchos los conceptos abstractos que se pretenden analizar y comprender en el área, y que por lo tanto suelen ser medidos indirectamente a través de los ítems que los integran.

Si se diseña un nuevo constructo de medición, la validación del mismo es el primer paso antes de realizar otros análisis estadísticos. Por ello, la validación es tan importante y ha de realizarse con suma rigurosidad, porque de ello depende la fiabilidad y validez de los resultados que se obtengan.

Cuando la valoración de los ítems se realiza a través de escalas ordinales se hace necesario utilizar una metodología que precisamente tenga en cuenta este hecho, es decir, tenga en cuenta las características de los datos recopilados. Sin embargo, habitualmente hay estudios en el área que comenten el error de relajar las condiciones necesarias para la medición, y trabajan con datos ordinales como si fueran variables de intervalo (Bond & Fox, 2007). En relación con este hecho, existen diversos estudios que analizan hasta qué punto es válido tratar las variables ordinales como si fueran continuas. Por ejemplo, el trabajo de Rhemtulla (2012) concluye que son necesarias al menos cinco categorías para que puedan utilizarse metodologías basadas en variables continuas sin que se produzcan diferencias significativas en los resultados obtenidos. Por otro lado, el trabajo de Salzberger y Sinkovics (2006) realiza una comparación entre el uso de la Teoría de la Medición de Rasch y el tradicional análisis factorial, destacando las bondades del primero sobre el segundo. Se observa por lo tanto que este es un tema de interés en el que no hay conclusiones contundentes y, por lo tanto, más estudios comparativos son necesarios.

La Teoría de la Medición de Rasch es una herramienta que pretende ofrecer la solución a este problema mediante la obtención de unidades de intervalo que cumplan los requisitos de medición a partir de datos categóricos. El uso de esta metodología no está muy extendido en el campo de operaciones, aunque cada vez son más los estudios que lo utilizan en el ámbito de la administración y gestión de empresas (Sanchez & Blanco, 2012).

En base a lo anterior, el objetivo de este estudio es presentar un caso de aplicación de la Teoría de la Medición de Rasch para la validación de constructos. Para ello, de forma detallada se describirán todas las fases a seguir a la hora de validar un constructo: dimensionalidad del constructo, fiabilidad y validez globales de las medidas, fiabilidad y validez individuales de las medidas de los ítems y de los sujetos, y análisis de las categorías de respuesta. Consideramos que el estudio puede ser de gran interés para otros investigadores que deseen aplicar esta metodología en el diseño de sus estudios.

Además, para completar el análisis se mostrará un ejemplo de validación con el constructo “Motivación para implantar la mejora continua”. Se comenzará presentando la revisión bibliográfica realizada. A continuación, se procede a diseñar un constructo a partir de la consulta a expertos. Después se realizarán las comprobaciones de validez antes citadas que permitirán concluir la definición de un constructo válido integrado por 10 ítems.

2. Marco Teórico

2.1. El problema de la medición en ciencias sociales

En el campo de las Ciencias Sociales en general y en el área de Administración de Empresas en particular, existen infinidad de realidades que no pueden medirse directamente como ocurre con multitud de magnitudes físicas como el peso, la longitud o la altura, entre otras. Ante esta circunstancia, la medición suele realizarse indirectamente mediante la valoración de un conjunto de ítems que, en teoría, integran el constructo o realidad latente en la que estamos interesados.

De acuerdo con la clasificación de Stevens (1959) en (Oreja, 2005), los números que se obtienen de esas mediciones pueden pertenecer a cuatro escalas diferentes: nominal, ordinal, de intervalo y ratio. En función de la escala en la que nos encontremos los cálculos estadísticos a aplicar son distintos.

- Nominal: los números equivalen a etiquetas.
Ejemplo: 0 = masculino; 1 =femenino.
- Ordinal: la asignación de números ayuda a establecer un orden.
Ejemplo: 1 = desacuerdo; 2 = neutral; 3= De acuerdo
- Intervalo: cumplen el principio de aditividad. La única exigencia es la existencia de un origen común. Es el tipo de valores que utiliza el Rasch.
- Ratio: similar a intervalo, pero con un punto de cero absoluto.

En el caso de las puntuaciones ordinales, que son la materia prima con la que trabaja el Rasch, las puntuaciones asignadas a los ítems pueden ser del tipo “Sí/No”, “Bien/Mal”, transformándose en observaciones $x= 0,1$, en el caso de variables dicotómicas; o pueden ser escalas de respuesta politómicas del tipo “Acuerdo/Desacuerdo/Completamente en desacuerdo”, las cuales se convierten en observaciones $x=1,2,3$ por ejemplo (Wright, Mok 2000).

En cualquiera de los dos casos estas puntuaciones no pueden ser consideradas mediciones ya que, para que un constructo sea medible ha de poseer una estructura aditiva. Por lo tanto, si se utiliza una escala Likert y se tratan sus puntuaciones como si fueran variables de intervalo, ignorando que las variables ordinales no tienen características aditivas, se estaría faltando al rigor científico exigido para una medición fundamental de la variable latente analizada, tal y como indicaron Bond y Fox (2007).

La principal consecuencia de este hecho es que, al no tratarse de medidas de intervalo, la aplicación de cualquier estadístico como la media o la mediana no tiene sentido y es incorrecto (Wright & Mok 2000).

Norman Campbell (1921/1953), miembro activo de la Asociación Británica para el Avance de las Ciencias, definió medir como: *la asignación de números a sistemas de acuerdo con las leyes científicas*. Campbell también destacó que, de acuerdo con su planteamiento metodológico, el principal aspecto que no cumplen las mediciones de los hechos sociales es la concatenación de medidas, base del principio de aditividad. Por ello, exigió que se demostrara la aditividad de las mediciones de las realidades sociales.

En respuesta a este requerimiento, a partir de los años 50 del siglo pasado, algunos investigadores sociales comenzaron a desarrollar estudios sobre medición. Es el caso de los trabajos de medición iniciados por Rasch (1960/80,1977) y continuados por otros (Wright, Andrich, Masters, Linacre) que trataron de cumplir los requisitos de medición planteados por Campbell, evidenciando la posibilidad de mediciones objetivas en las ciencias sociales.

2.2. La Teoría de la Medición de Rasch

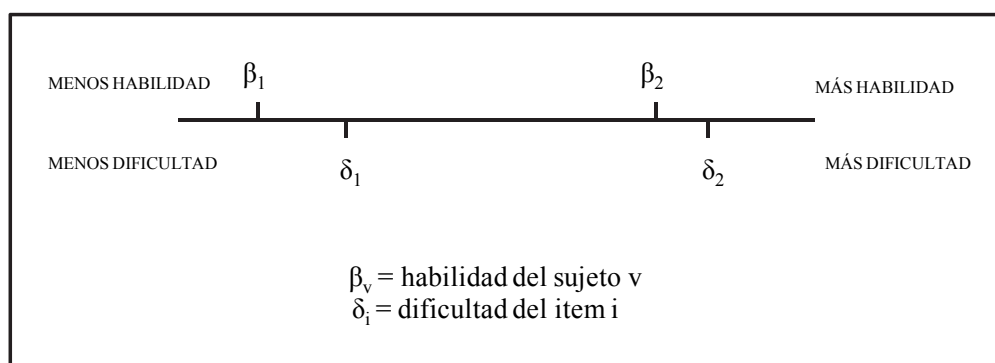
La Teoría de la Medición de Rasch se enmarca dentro de la Teoría Moderna del Test o Teoría del Rasgo Latente. A diferencia de su predecesora, la Teoría Clásica del Test, ésta es capaz de garantizar el principio de invarianza y, por lo tanto, los resultados no dependen de la muestra de sujetos ni de ítems utilizados.

La Teoría Moderna del Test engloba dos tipos de modelos: los modelos de contraste y los modelos de ajuste. La diferencia entre ambos se refleja en el tipo de relación que se establece entre los datos y el modelo estadístico utilizado. Así, los modelos de contraste (Teoría de la Respuesta al Ítem, TRI), buscan el modelo que mejor se ajuste a los datos mediante la parametrización del rasgo latente y de las propiedades de los ítems. Mientras que en los modelos de ajuste, entre los que se encuentra la Teoría de la Medición de Rasch, se parte de un modelo determinado que dispone de unas propiedades específicas a las que los datos obtenidos se ajustan o no. Del análisis de los desajustes se obtiene la información necesaria para determinar las calibraciones de ítems y las medidas convenientes en cada caso (Oreja-Rodríguez, 2008).

La Teoría de la Medición de Rasch se caracteriza por cumplir los siguientes principios:

- Unidimensionalidad: se consigue cuando todos los ítems (preguntas del cuestionario) se refieren a un único constructo o rasgo latente, de tal manera que todos ellos pueden ubicarse en un mismo continuo lineal junto con los sujetos (Ilustración 1).

Ilustración 1. Continuo lineal



Fuente: Sánchez y Blanco (2012)

- Invarianza (objetividad específica): garantiza que los resultados obtenidos del análisis no dependen de la muestra de sujetos ni de los ítems utilizados.
- Aditividad: la distancia entre los valores 3 y 4, por ejemplo, es la misma que la distancia entre los valores 4 y 5. Esto es una característica que no cumplen las mediciones ordinales pero sí las de intervalo.

Adicionalmente, han de destacarse dos características que resaltan la riqueza de la metodología. La primera de ellas es la suficiencia estadística, entendida ésta como la capacidad para extraer conclusiones fiables a partir de muestras pequeñas. Y la segunda es la capacidad de identificar y analizar los casos anómalos. Esto no es posible con otras técnicas donde los casos anómalos son eliminados en el análisis por considerar que desvirtúan la muestra. La Teoría de la Medición de Rasch, sin embargo, los identifica para que el investigador pueda analizar las razones que lo convierten en un caso “raro”. Ha de tenerse en

cuenta que, muchas veces, obtenemos más información de este tipo de casos que de los clasificados como casos “normales”, especialmente en el ámbito empresarial.

A la hora de realizar la transformación de variables ordinales a variables de intervalo, el Modelo se basa en la siguiente afirmación: un sujeto con mayor habilidad que otro deberá tener una mayor probabilidad de resolver un ítem correctamente (contestar afirmativamente o con mayor puntuación en la escala planteada). Igualmente, ante un ítem más difícil que otro, la probabilidad de que cualquier sujeto conteste correctamente al fácil es mayor que la probabilidad de que conteste correctamente al difícil (Rasch, 1960) en (Bond & Fox, 2007).

Así, en la Ilustración 1, el sujeto 2 tiene más probabilidad que el sujeto 1 de responder correctamente a cualquier de los ítems. Igualmente, para cualquiera de los sujetos, la probabilidad de responder correctamente al ítem 1 es mayor que la de responder correctamente al ítem 2.

Por lo tanto, el modelo matemático se deriva de la función logística que relaciona la probabilidad creciente de respuesta a los ítems con la habilidad de los sujetos (González Montesinos, 2008).

A partir de este principio, se determina un concepto adicional: la distancia. Ésta se entiende como la diferencia entre la habilidad del sujeto (β) y la dificultad del ítem (δ). Una vez definido el concepto, se pueden establecer tres situaciones:

- Si $\beta - \delta > 0$ entonces la probabilidad de que el sujeto conteste correctamente a un ítem será más del 50%
- Si $\beta - \delta < 0$ entonces la probabilidad de que el sujeto conteste correctamente a un ítem será menos del 50%
- Si $\beta - \delta = 0$ entonces la probabilidad de que el sujeto conteste correctamente a un ítem será del 50%

Partiendo de esta idea se desarrolla la expresión matemática del Modelo, que puede ser consultada en cualquiera de los manuales ya existentes sobre la metodología (ver Bond & Fox, 2007; Alagumalai, Curtis & Hnugi, 2005; von Davier & Carstensen, 2007).

En lo referente al tratamiento de los datos, la Teoría de la Medición de Rasch cuenta ya con varios paquetes informáticos que facilitan su uso (Facets, Winsteps...). Algunos de ellos son de uso libre como el Ministeps (<http://www.winsteps.com/ministep.htm>) (Linacre, 2012a) que cuenta con las utilidades reducidas respecto al programa completo ya que únicamente permite trabajar con 25 ítems y 75 sujetos. En el presente estudio se ha utilizado la versión 3.75 del Winsteps. Por otro lado, cabe destacar que otros software como R y SPSS ya han desarrollado paquetes estadísticos para aplicar la metodología de Rasch. Sin embargo, los autores desconocen la potencialidad de esos paquetes informáticos y, por ello, quedaría pendiente un estudio comparativo entre el uso de los distintos software.

2.3. Aplicaciones de la metodología

El uso de la Teoría de la Medición de Rasch permite al investigador realizar múltiples análisis y comprobaciones. Entre ellas podemos enunciar las siguientes: cálculo de la fiabilidad y de la validez globales de las medidas; fiabilidad de los sujetos y de los ítems de forma individual; comprobación de la hipótesis de unidimensionalidad; análisis de las categorías de respuesta; jerarquización de los ítems en función de su dificultad y de los sujetos en función de su habilidad; detección de diferencias significativas entre sujetos de distintos grupos (para ello previamente habrá de haberse fijado una variable de diferenciación como

puede ser el género de los sujetos, el tamaño si hablamos de empresas, la situación por Comunidades, etc.). Además, en las últimas versiones, se incluye una función a través de la cual se pueden detectar puntos fuertes y puntos débiles (una especie de análisis DAFO) de cada sujeto de forma individual.

En las siguientes líneas se hará referencia a algunas de esas aplicaciones, concretamente a aquellas que son necesarias para la validación de constructos. El objetivo es clarificar el análisis a realizar, así como ofrecer los valores óptimos o de referencia que han de tenerse en cuenta.

2.3.1. Dimensionalidad del constructo

En apartados anteriores de este mismo capítulo, se han explicado los principios que subyacen tras la Teoría de la Medición de Rasch. Es condición necesaria para poder hacer uso de esta metodología comprobar que el constructo definido es unidimensional. Además de por motivos metodológicos, esta comprobación es interesante cuando se están definiendo nuevos constructos, ya que permite confirmar la correcta definición de los mismos. Este aspecto adquiere vital importancia en la presente investigación.

Para el análisis de la dimensionalidad, la Teoría de la Medición de Rasch hace uso del Análisis de los Componentes Principales. Este análisis busca patrones en los residuales de los datos que no se ajusten al Modelo de Rasch (primera dimensión o dimensión principal). Es decir, analizan la parte “inesperada” de los residuales de los datos. En concreto, se intenta localizar si existen grupos de ítems que compartan un atributo inesperado. Este atributo sería la “segunda dimensión” y vendría indicado por el primer contraste (Linacre, 2012b).

El Análisis de Componentes Principales tiene como objetivo principal rechazar la hipótesis de que “los residuales son ruido aleatorio” y encontrar el componente que explica el mayor porcentaje de la varianza de los mismos (Linacre, 2012b).

La detección de la posible segunda dimensión se lleva a cabo mediante el análisis de los datos que aparecen en la tabla 23 del Winsteps (Variance components screen plot for ítems) que muestra la descomposición de la varianza de las observaciones de los ítems (Linacre, 2012a). A continuación se ofrece la salida adaptada de dicha tabla (Tabla 1).

Tabla 1. Tabla de los residuales estandarizados de la varianza (Autovalores)

	Empírico			Modelo
Varianza total en los valores observados	20,4	100,0%		100,0%
Varianza explicada por las medidas	10,4	51,1%		51,3%
Varianza explicada por los sujetos	1,9	9,4%		9,5%
Varianza explicada por los ítems	8,5	41,6%		41,8%
Varianza no explicada total	10,0	48,9%	100,0%	48,7%
Varianza no explicada en el 1º factor	1,9	9,1%	18,7%	

Fuente: Elaboración Propia

El significado de cada uno de los apartados es el siguiente:

- Varianza explicada por las medidas: varianza en las observaciones explicada por las medidas de Rasch
- Varianza explicada por los sujetos: varianza explicada por la habilidad de las personas
- Varianza explicada por los ítems: varianza explicada por la dificultad de los ítems
- Varianza total no explicada: varianza en las observaciones no explicada por las medidas de Rasch
- Varianza no explicada en el factor: varianza no explicada por las medidas de Rasch que sí explica el contraste o componente.

Se observa que, para cada uno de los apartados, aparecen dos valores, uno empírico y uno modelizado. El primero de ellos se refiere a los componentes de la varianza para los datos observados. El segundo, por el contrario, es la composición de la varianza si los datos se ajustaran perfectamente al Modelo de Rasch.

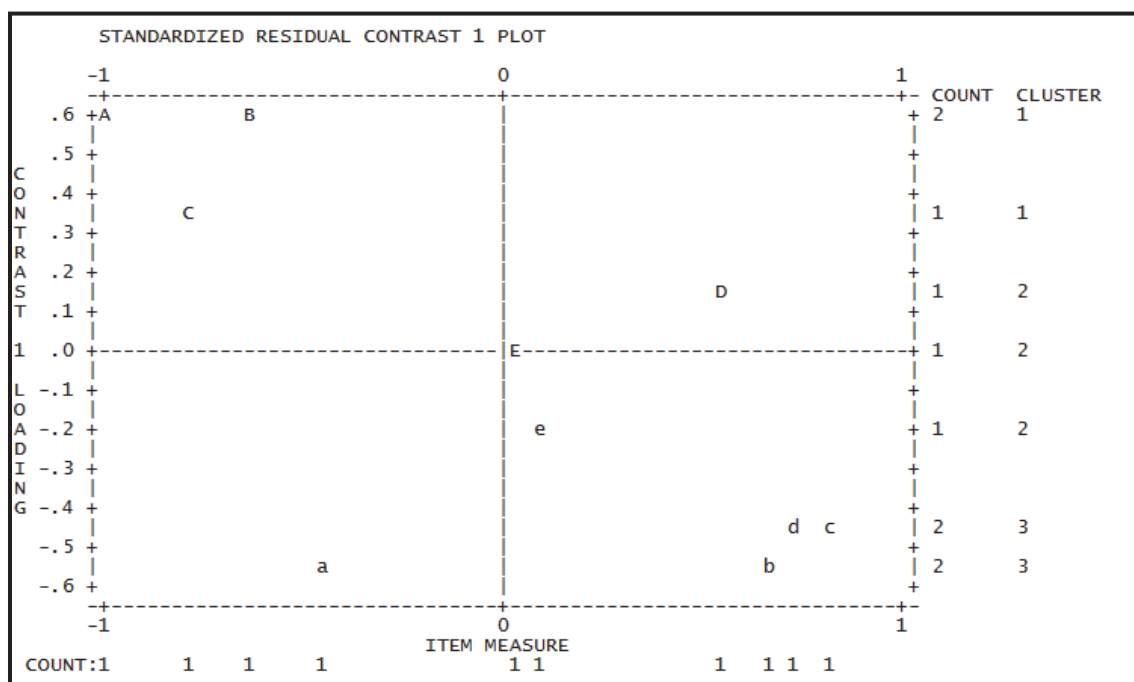
Partiendo del hecho de que la unidimensionalidad no es perfecta, es necesario establecer unos límites o valores objetivos que nos digan cuándo la segunda dimensión detectada es lo suficientemente fuerte para distorsionar las mediciones. En el manual del programa informático se ofrecen unas guías para la interpretación de los resultados de carácter orientativo.

De acuerdo con Linacre (2012b) los valores de referencia que han de cumplirse para garantizar la unidimensionalidad de un constructo son:

- El autovalor del primer contraste ha de ser menor o igual a dos. Siguiendo la hipótesis mencionada anteriormente, un autovalor menor que dos no permite rechazar la hipótesis “los residuales son ruido aleatorio”.
- El porcentaje de varianza explicada por el primer contraste ha de ser menor que el porcentaje de varianza explicado por los ítems.

Con todo, si una vez realizado el análisis se detecta la presencia de una segunda dimensión, se procederá a identificar los ítems que podrían estar integrándola. Para ello, se acude al gráfico 23.2 del Winsteps. En el citado gráfico aparecen representados los residuales (ver Ilustración 2). Los residuales que más se alejen del resto, verticalmente hablando, nos están indicando qué ítems podrían ser los causantes de la segunda dimensión. En el ejemplo de la Ilustración 2 serían los ítems A y B principalmente. Una vez identificados los ítems, es labor del investigador decidir en base a la información que posee si es oportuno o no eliminar el ítem del constructo.

Ilustración 2. Gráfico de cargas de los residuales



Fuente: Elaboración Propia

2.3.2. Fiabilidad y validez de las medidas

En relación a la validez de las medidas, se dice que una medida es válida cuando mide lo que se supone que ha de medir. El análisis global de la validez se realiza a través de la tabla 3.1. del Winsteps. Concretamente a través de los índices INFIT y OUTFIT (Tabla 2).

Tabla 2. Fiabilidad y validez globales de las medidas

	INFIT		OUTFIT		FIABILIDAD	CORRELACIÓN
	MNSQ	ZSTD	MNSQ	ZSTD		
Sujetos	1,03	-0,1	1,08	0	0,64	0,95
Ítems	1,00	-0,1	1,08	0,1	0,98	-1,00

Fuente: Elaboración Propia

El índice INFIT es un estadístico de ajuste más sensible a los valores inesperados próximos al nivel del sujeto/ítem; mientras que el índice OUTFIT es un estadístico más sensible a los valores alejados del nivel del sujeto/ítem.

Cada uno de los estadísticos se expresa a través de dos datos (Linacre, 2002):

- Media cuadrática no estandarizada (MNSQ): Es una transformación de los residuales generada por la diferencia entre la observación y la estimación del Modelo de Rasch. Se estiman mediante

estadísticos Chi-cuadrados divididos por sus grados de libertad. El valor esperado óptimo es 1, pero se admiten como aceptables aquéllos situados en el rango entre 0,5 y 1,5 (Linacre, 2002; Linacre, 2012b; Oreja, 2005; Febles Acosta, 2008). Concretamente:

- Valores inferiores a 0,5: evidencian desajustes, pero no distorsionan.
 - Valores entre 1,5 y 2: indican desajuste, son improductivos pero utilizables.
 - Valores superiores a 2: generan distorsiones en la medición.
- Varianza normalizada (ZSTD): Se presentan como valores normalizados z , con una distribución con media 0 y varianza 1. Usualmente se obtienen mediante la conversión del estadístico media cuadrática en una z -normalizada normalmente distribuida por medio de la transformación de Wilson-Hiferty (Linacre, 2002). Su valor óptimo es cero (Linacre, 2012; Oreja, 2005; Febles Acosta, 2008). Sin embargo:
 - Se admiten como aceptables aquellos que varíen entre -1,9 y 1,9.
 - Valores superiores a 2 están desajustados e indican mayor variabilidad en los datos que lo que predice el Modelo de Rasch.
 - Valores inferiores a (-2) están sobre ajustados, evidencian dependencia en los datos.

Al analizar la validez, lo primero que hay que detectar son desviaciones en las medias cuadráticas, especialmente aquellas con valores superiores a 2. Si por el contrario las desviaciones se detectan en las varianzas estandarizadas (ZSTD), estas pueden ignorarse siempre y cuando el valor de MNSQ sea aceptable (Linacre, 2012b).

A la hora de analizar la fiabilidad de las medidas, el modelo de Rasch utiliza índices de separación de sujetos e ítems. El índice de separación de los sujetos estima la repetición de la ubicación de los sujetos a lo largo del continuo lineal si se utiliza otra muestra de ítems que mida ese mismo constructo. Igualmente, el índice de separación de los ítems estima la repetición de la ubicación de los ítems a lo largo del continuo lineal si se utiliza otra muestra de sujetos (Oreja, 2005). Cuanto mayor sea el valor del índice, mejor es la separación, más precisa la medición, luego más alta la fiabilidad de separación de Rasch. Con todo, cabe destacar que el valor óptimo para la fiabilidad de separación de Rasch es de 1 tanto para sujetos como para ítems (Linacre, 2012b).

En lo que se refiere a la correlación, ésta es la correlación de Pearson (incluye puntuaciones extremas) y se espera que sea 1 para sujetos y -1 para ítems cuando la información o los datos están completos (Linacre, 2012b).

Para comprobar la fiabilidad global y la correlación también hemos de acudir a la tabla 3.1. del Winstep (ver dos columnas de la derecha de la Tabla 2).

Además del análisis global de la validez de las medidas, la metodología permite realizar un análisis de la validez individual de las medidas de cada ítem y de cada sujeto. Dicho análisis se realiza también a través de los índices INFIT y OUTFIT, y los valores óptimos y aceptables son los mismos que en el caso global.

Cuando las medidas de un sujeto o ítem presentan desajustes, el modelo permite identificar en dónde se producen dichos desajustes mediante la tabla ofrecida por el Winsteps denominada “person misfitting responses” en el caso de los sujetos y la tabla “Item misfitting responses” del Winsteps en el caso de los ítems. En la Ilustración 3, se ofrece un fragmento de la tabla “Item misfitting responses”. En la citada Ilustración, se observa que el sujeto 73 presenta desajustes en los ítems 2 y 3. Sabemos que hay desajustes porque, debajo de la puntuación que este sujeto ha dado a los ítems 2 y 3, en este caso los puntuó con un 5

a los dos, aparece un número. Cuando ese número es positivo, quiere decir que ese sujeto sobrevaloró ese ítem y cuando es negativo significa que el sujeto infravaloró el ítem en cuestión. Para determinar el desajuste el modelo se basa en el resto de respuestas del sujeto y en las respuestas de otros sujetos a ese mismo ítem. En el ejemplo el sujeto ha sobrevalorado ambos ítems.

Una vez detectado el desajuste, es labor del investigador analizar las posibles causas del mismo. En ocasiones, el análisis de los desajustes aporta información verdaderamente valiosa.

Ilustración 3. Tabla de desajustes del sujeto

TABLE OF POORLY FITTING PERSON (ITEM IN ENTRY ORDER)									
NUMBER	NAME	MEASURE - INFIT (MNSQ) OUTFIT							
73	E217	1	2	1	1	1	4	8	1
OBSERVED:	1:	1	5	5	1	1	1	1	1
Z-RESIDUAL:		7	2						

Fuente: Elaboración Propia

2.3.3. Análisis de categorías

El objetivo de este análisis es comprobar que los datos de la muestra analizada se ajustan a las categorías de respuesta definidas. Para ello acudimos a la información aportada por la tabla 3.2. del Winsteps cuyos datos principales aparecen recogidos en la Tabla 3.

Tabla 3. Análisis de categorías

Categoría	Observado		Media	Media	INFIT	OUTFIT	Umbral	Medida	
Nombre	Marca	Frecuencia	%	observada	esperada	MNSQ	MNSQ	de AN-DRICH	de las categorías
1	1	440	46	-0.97	-0.96	0.96	0.94	NONE	(-1.38)
2	2	111	12	-0.59	-0.57	0.87	0.99	0.59	-0.53
3	3	110	11	-0.12	-0.20	0.83	0.64	-0.37	-0.04
4	4	119	12	0.23	0.15	0.79	0.72	-0.10	0.48
5	5	180	19	0.37	0.43	1.23	1.72	-0.12	(1.49)

Fuente: Elaboración Propia

El significado de las diferentes cabeceras de la tabla es el siguiente:

- Nombre de la categoría: número de la categoría tras la codificación.
- Puntuación de la categoría: Valor de la categoría
- Cómputo observado: Valor absoluto y porcentaje del número de casos en la categoría.
- Media observada: La “medida media” (average measure) de una categoría es la habilidad media de la persona que responde a esa categoría. Es un valor real, no es un parámetro del Modelo de Rasch.

En teoría los sujetos que escogen las categorías más altas de respuesta tienen medias más altas. Por esto, se espera que el valor de la media observada aumente con la categoría de respuesta.

- Media esperada: es el valor esperado de la media para esta categoría. En este caso los valores siempre son crecientes.
- INFIT MNSQ: es la media de las medias estandarizadas del INFIT para las respuestas en cada categoría. Tiene un valor esperado de 1.
- OUTFIT MNSQ: es la media de las medias estandarizadas del OUTFIT para las respuestas en cada categoría. Tiene un valor esperado de 1.
- Umbral de Andrich: es la medida calibrada de la transición de una categoría a la siguiente. La primera categoría no tiene transición, por eso aparece NONE. El valor de los umbrales se espera que aumente a medida que aumenta la categoría de respuesta en la que nos encontramos. Si esto no se cumple y hay un desorden, éste nos puede estar indicando que una determinada categoría no es muy utilizada y que una redefinición de las categorías podría ser necesaria.
- Medida de la categoría: es la medida de esa categoría independiente de la muestra utilizada.

Dicho esto, siguiendo las recomendaciones del manual (Linacre, 2012b), afirmaremos que las categorías definidas son correctas cuando se cumple lo siguiente:

- Más de 10 observaciones por categoría
- Una distribución similar entre categorías
- Que las “medidas medias” crezcan.
- Los valores de las “medidas medias” similares a los esperados.
- Validez de las observaciones en las diferentes categorías: valores de OUTFIT MNSQ cercanos a 1.

3. Metodología

A lo largo de este apartado se procederá a explicar las diferentes etapas que componen el proceso de diseño de un constructo. El apartado de validación a partir de la Teoría de la Medición de Rasch, objeto principal de este trabajo, se incluye en el apartado de resultados.

Tal y como se dijo en el apartado introductorio, la explicación del proceso de validación se realizará mediante el uso de un ejemplo, en concreto, vamos a validar el constructo “motivación para implantar la mejora continua”. Mediante la validación de este constructo, además, se avanzará en la teoría de la mejora continua. Este campo, pese a contar con una amplia trayectoria, necesita validar y sistematizar la teoría establecida en estudios previos (Marín-García & García-Sabater, 2010).

3.1. Revisión bibliográfica

En primer lugar, los autores procedieron a realizar una búsqueda bibliográfica sobre el tema objeto de estudio. Para realizar la búsqueda se buscaron artículos científicos sobre el tema en las bases de datos Web of Science, Scopus y Dialnet.

Se observó que, pese a la amplia literatura existente sobre el tema de la mejora continua, no existía ningún constructo definido y validado que permitiera analizar la motivación de las empresas para implantar la mejora continua. Sí existían estudios que, de forma independiente, analizaban las motivaciones de las

empresas pertenecientes a una muestra concreta. Sin embargo, no existía un constructo definido y validado que permitiera ampliar la teoría y ser utilizado en más estudios de forma estandarizada.

Una vez detectada la necesidad, también a través de la revisión bibliográfica, se identificaron los motivos para implantar la mejora continua que aparecían a lo largo de la literatura académica. En la tabla 4, se recoge un resumen con los distintos motivos y los autores que lo definían.

Tabla 4. Resumen de razones identificadas en la revisión bibliográfica

RAZONES O MOTIVOS	AUTORES
Añadir valor al cliente / Enfoque al cliente	(Hung, 2006); (Bullita, 2006) en (Kohlbacher, 2010); (Classe & Mundle, 1997) en (Kohlbacher, 2010); (Küng & Hagen, 2007); (Furey, 1993); (Silvestro & Westley, 2002); (Clemmer, 1994); (Paper & Chang, 2005); (Lee & Dale, 1998); (Hill & Collins, 1998); (Corallo et al., 2010); (Melan, 1989); (Armistead & Pritchard, 1999); (Biazzo & Bernardi, 2003); (Sandhu & Gunasekaran, 2004)
Aumento de la competencia nacional y/o internacional	(Hung, 2006); (Classe & Mundle, 1997) en (Kohlbacher, 2010); (Küng & Hagen, 2007); (Feltes & Karuppan, 1995); (Armistead & Machin, 1998); (Clemmer, 1994); (Hill & Collins, 1998); (Ongaro, 2004)
Necesidad de ahorrar dinero mediante la reducción de costes o el aumento de la productividad	(Paper & Chang, 2005); (Clemmer, 1994); (Davenport & Short, 1990); (Palmberg, 2010); (Hill & Collins, 1998); (Corallo et al., 2010)
Como parte de un proyecto más amplio (sistema de calidad total, ISO 9000; adopción modelo de calidad, etc.)	(Paper & Chang, 2005); (Pauls, 2006) en (Kohlbacher, 2010); (Armistead & Machin, 1998); (Corallo et al., 2010)
La empresa no era capaz de hacer frente al crecimiento	(Pauls, 2006) en (Kohlbacher, 2010)
Tasa de fracaso de los productos muy elevada/ mejorar la calidad	(Wahlich, 2004) en (Kohlbacher, 2010); (Davenport & Short, 1990)
Necesidad de mejorar los productos existentes, crear nuevos o entrar en nuevas líneas de negocio para continuar siendo competitivos	(Paper & Chang, 2005); (Melan, 1989); (Sandhu & Gunasekaran, 2004); (McAdam & McCormack, 2001)
Mejorar la efectividad organizacional	(Jones, 1994) en (Palmberg, 2010); (Elzinga et al., 1995); (Armistead & Pritchard, 1999)
Eliminar las barreras existentes entre los departamentos funcionales	(Paper & Chang, 2005); (Jones, 1994) en (Palmberg, 2010); (Llewellyn & Armistead, 2000)
Necesidad de mejorar la gestión de los recursos de IT	(Paper & Chang, 2005); (Sandhu & Gunasekaran, 2004); (Hill & Collins, 1998); (Ongaro, 2004)
Motivación interna de la propia empresa	(Mittermaier & Braun, 2004) en (Kohlbacher, 2010); (Clemmer, 1994)
Una crisis obligó a analizar las prácticas que se estaban llevando a cabo	(GropengieBer, 1997) en (Kohlbacher, 2010)

Conseguir reducciones de tiempo	(Davenport & Short, 1990); (Ongaro, 2004)
Para identificar oportunidades para subcontratar	(Lindsay, Downs & Lunn 2003, Lock Lee 2005)
Para identificar oportunidades para utilizar la tecnología apoyando el negocio	(Lindsay, Downs & Lunn 2003, Lock Lee 2005)
Aumentar la calidad del aprendizaje colectivo dentro de la organización y entre la organización y su entorno	(Bawden & Zuber-Skerritt, 2002)
Aumentar la comprensión de la estrategia y de las necesidades de los clientes entre los empleados	(Palmberg, 2010)
Estandarización de los procesos	(Palmberg, 2010)
Como parte de una estrategia de diferenciación	(Silvestro & Westley, 2002)
Debido a un evento importante (fusión o adquisición)	(Paper & Chang, 2005)
Presión por parte de los proveedores	Propio

Fuente: Elaboración Propia

3.2. Validación mediante grupo de expertos

El número de motivaciones encontrado fue tal que, de cara a garantizar la utilidad y la fiabilidad del mismo, se procedió a validar su contenido mediante una consulta a expertos. Un total de ocho expertos fueron consultados. Estos pertenecían al mundo de la empresa o a la academia. Así, entre ellos había Catedráticos de Administración de Empresas, gerentes de empresas de manufactura y servicios, responsables de calidad y consultores de calidad.

Los expertos, a partir del listado de motivaciones de la tabla 4, procedieron a dar su opinión sobre la idoneidad o no de incluir el ítem en el constructo. Sus principales aportaciones, además de la inclusión o no de un determinado ítem, eran sugerencias de agrupación de ítems bajo un único factor dado su similitud o sugerencias de redacción.

Finalmente, tras la consulta a expertos, el constructo “motivación para implantar la mejora continua” quedó integrado por diez ítems (tabla 5).

Tabla 5. Constructo “Motivación para implantar la mejora continua”

MC3.1	Conocer lo que está ocurriendo en la empresa
MC3.2	Presión de los clientes
MC3.3	Auditar la cultura de la organización
MC3.4	Medición del desempeño
MC3.5	Identificar oportunidades de mejora
MC3.6	Presión de los proveedores
MC3.7	Benchmarking interno
MC3.8	Como parte de un sistema de políticas retributivas
MC3.9	Como parte de un sistema más amplio (Lean Management, ISO 9000...)
MC3.10	Conocer lo que quieren los clientes

Fuente: Elaboración Propia

3.3. Envío de la encuesta a empresas

El constructo definido formaba parte de una encuesta más amplia sobre distintos factores de la mejora continua (motivaciones, obstáculos, facilitadores, beneficios, etc.). La población del estudio estaba integrada por las empresas de la Comunidad Autónoma de Cantabria que practicaran la mejora continua.

En primer lugar, se sondeó a todas las empresas de Cantabria con más de 20 empleados (808 empresas) para saber si practicaban o no la mejora continua. Un total de 299 empresas respondieron (37% de respuesta), afirmando 209 practicar la mejora continua. A esas 209 empresas se les envió la encuesta completa incluyendo el constructo “motivación para implantar la mejora continua”.

En concreto, las empresas tenían que valorar en una escala Likert de 1 a 5 en qué medida cada uno de los ítems había sido una motivación o no para ellas cuando decidieron implantar la mejora continua. Así, elegirían 1 si no tuvieron en cuenta ese motivo y 5 si fue una de las razones principales que les llevó a implantar la mejora continua

Finalmente, 109 empresas respondieron a la encuesta. La validación del constructo y otros análisis posteriores que no son objeto de este trabajo se realizan a partir de esos datos.

4. Resultados

Una vez obtenidas las respuestas de las 109 empresas, en este apartado vamos a mostrar las distintas etapas a seguir para validar un constructo con la Teoría de la Medición de Rasch. En concreto, las pruebas a realizar son las siguientes:

- A) análisis de la dimensionalidad del constructo;
- B) fiabilidad y validez globales de las medidas;
- C) fiabilidad y validez individuales de los sujetos;
- D) fiabilidad y validez individuales de los ítems; y
- E) análisis de las categorías de respuesta.

Cabe destacar que cualquier cambio en el número de sujetos o de ítems incluido en la muestra derivado de alguna de las comprobaciones, supone tener que comenzar de nuevo con todas las pruebas.

A.1) ANÁLISIS DE LA DIMENSIONALIDAD: CASO 1 (109 SUJETOS Y 10 ÍTEMS)

La primera comprobación a realizar es el análisis de la dimensionalidad a partir de la información recogida en la tabla 6.

Tabla 6. Tabla de los residuales estandarizados de la varianza (Caso 1)

	Empírico			Modelo
Varianza total en los valores observados	20,5	100,0%		100,0%
Varianza explicada por las medidas	10,5	51,2%		51,1%
Varianza explicada por los sujetos	3,0	14,5%		14,4%
Varianza explicada por los ítems	7,5	36,8%		36,7%
Varianza no explicada total	10,0	48,8%	100,0%	48,9%
Varianza no explicada en el 1º factor	1,7	8,4%	17,2%	

Fuente: Elaboración Propia

Los resultados obtenidos nos permiten afirmar que el constructo definido es unidimensional ya que el autovalor del primer contraste es menor de dos y el porcentaje de varianza explicada por el primer factor es menor que la varianza explicada por los ítems.

B.1) FIABILIDAD Y VALIDEZ GLOBALES DE LAS MEDIDAS: CASO 1 (109 SUJETOS Y 10 ITEMS)

En lo referido a la fiabilidad y la validez globales de las medidas (Tabla 7), podemos concluir para este apartado que la validez es muy buena ya que todos los valores, tanto los de INFIT como OUTFIT, se aproximan a los óptimos. En el caso de la fiabilidad, concluimos que ésta es muy buena para los ítems y aceptable para los sujetos. En lo referido a la correlación es excelente en ambos casos.

Tabla 7. Fiabilidad y validez globales de las medidas (Caso 1)

	INFIT		OUTFIT		FIABILIDAD	CORRELACIÓN
	MNSQ	ZSTD	MNSQ	ZSTD		
Sujetos	1,03	-0,1	1,11	0,0	0,77	0,99
Ítems	1,02	-0,1	1,11	0,3	0,98	-1,00

Fuente: Elaboración Propia

C.1) FIABILIDAD Y VALIDEZ INDIVIDUALES DE LOS SUJETOS: CASO 1 (109 SUJETOS Y 10 ITEMS)

De acuerdo con la información recogida en la Ilustración 4, hay un total de 17 sujetos que presentan desajustes. Estos sujetos son los siguientes: 22, 94, 36, 67, 31, 70, 52, 16, 53, 40, 23, 59, 28, 105, 86, 92 y 21.

Ilustración 4. Sujetos ordenados según desajuste que presentan (Caso 1)

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S. E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXP.	EXACT OBS%	MATCH EXP%	PERSON	
22	19	10	-1.04	.33	3.11	3.1	7.34	4.6	A-	-.09	.53	40.0	46.1	E065 3 2 2 2 1 4 5 1 1 1 3 1 2
94	35	10	-.38	.30	1.81	1.7	3.39	3.4	B-	-.15	.61	40.0	34.3	E249 2 2 1 1 1 3 4 1 1 1 3 1 2
36	26	10	-.39	.29	2.06	2.1	3.18	3.2	C	.04	.62	20.0	36.5	E102 1 1 1 1 1 3 2 1 1 1 1 1 2
67	16	10	-1.41	.38	1.35	.8	3.08	2.0	D	.08	.46	50.0	57.5	E181 3 2 2 2 1 2 3 4 1 1 3 3 3
31	16	10	-1.41	.38	1.28	.7	3.08	2.0	E	.14	.46	40.0	57.5	E095 2 2 2 2 1 1 3 2 1 1 3 1 4 1
70	37	10	-.57	.31	1.52	1.2	2.76	2.6	F	.28	.59	40.0	34.1	E190 2 1 1 1 1 1 4 4 1 1 2 3 4 1
52	14	10	-1.75	.45	2.53	1.8	1.32	.6	G	.38	.38	70.0	65.0	E133 1 2 2 2 1 1 4 4 1 1 3 1 1 2
16	23	10	-.65	.30	2.34	2.5	2.52	2.3	H	.33	.60	40.0	36.5	E047 2 2 2 2 2 1 4 8 1 1 3 3 2 4
53	38	10	-.67	.32	2.26	2.3	2.51	2.3	I	.00	.58	20.0	38.9	E135 3 2 2 2 2 1 4 2 1 1 3 3 1 1
40	22	10	-.74	.31	2.20	2.3	2.45	2.1	J	.24	.58	10.0	36.7	E109 3 2 2 1 1 1 4 8 1 1 3 1 4
23	36	10	-.47	.31	1.98	1.9	2.28	2.1	K	.41	.60	20.0	34.3	E066 3 2 2 2 1 1 4 8 1 1 2 3 2 2
59	31	10	.03	.29	1.90	1.9	2.26	2.3	L	.22	.63	30.0	29.3	E150 2 1 1 1 1 1 3 6 1 1 2 3 1 1
28	27	10	-.31	.29	2.22	2.4	2.19	2.1	M	.32	.63	40.0	36.3	E088 2 2 2 2 2 1 4 3 1 1 2 1 4 1
105	25	10	-.48	.29	1.97	2.0	2.20	2.0	N	.31	.62	20.0	36.9	E281 2 2 2 2 1 1 4 4 1 1 3 3 1 1
86	28	10	-.22	.29	1.74	1.6	2.16	2.1	O	.42	.63	20.0	27.8	E230 3 1 1 1 1 1 3 4 1 1 1 3 1 3
92	25	10	-.48	.29	1.99	2.0	2.04	1.8	P	.40	.62	30.0	36.9	E246 3 2 2 2 2 1 4 4 1 1 2 3 3 1
21	15	10	-1.57	.41	2.02	1.5	1.09	.4	Q	.43	.42	70.0	58.7	E061 3 1 1 1 1 1 3 1 1 1 3 2 1 2

Fuente: Elaboración Propia

Puesto que los desajustes podrían causar problemas en la fiabilidad y validez de las medidas, decidimos eliminarlos de la muestra en base al principio de invarianza. Sin embargo, tampoco hemos de olvidar la riqueza de la metodología y su capacidad para analizar casos anómalos.

A.2) ANÁLISIS DE LA DIMENSIONALIDAD: CASO 2 (92 SUJETOS Y 10 ITEMS)

Puesto que varios sujetos se han eliminado de la muestra, hemos de comenzar de nuevo las comprobaciones pero esta vez con 92 sujetos en vez de con 109. En la Tabla 8 se recoge la información objeto de análisis para determinar si el constructo, dadas las nuevas características de la muestra, es o no unidimensional.

Puesto que el autovalor del primer contraste es menor que 2 y el porcentaje explicado por los ítems es mayor al porcentaje explicado por el primer factor, se puede concluir que el constructo definido es unidimensional.

Tabla 8. Tabla de los residuales estandarizados de la varianza (Caso 2)

	Empírico			Modelo
Varianza total en los valores observados	20,5	100,0%		100,0%
Varianza explicada por las medidas	10,5	51,2%		51,1%
Varianza explicada por los sujetos	3,0	14,5%		14,4%
Varianza explicada por los ítems	7,5	36,8%		36,7%
Varianza no explicada total	10,0	48,8%	100,0%	48,9%
Varianza no explicada en el 1º factor	1,9	9,4%	19,3%	

Fuente: Elaboración Propia

B.2) FIABILIDAD Y VALIDEZ GLOBALES DE LAS MEDIDAS: CASO 2 (92 SUJETOS Y 10 ITEMS)

En lo referido a la fiabilidad y la validez globales de las medidas (Tabla 9), podemos concluir que la validez es muy buena ya que todos los valores, tanto los de INFIT como OUTFIT, se aproximan a los óptimos. En el caso de la fiabilidad, concluimos que ésta es muy buena para los ítems y aceptable para los sujetos. En lo referido a la correlación es excelente en ambos casos. Luego, se observa que apenas ha habido cambios respecto a la situación anterior, salvo un ligero empeoramiento en la validez de los sujetos que, no obstante, sigue dentro de los valores aceptables.

Tabla 9. Fiabilidad y validez globales de las medidas (Caso 2)

	INFIT		OUTFIT		FIABILIDAD	CORRELACIÓN
	MNSQ	ZSTD	MNSQ	ZSTD		
Sujetos	0,85	-0,5	0,82	-0,4	0,75	0,99
Ítems	1,02	-0,1	1,11	0,3	0,98	-1,00

Fuente: Elaboración Propia

C.2) FIABILIDAD Y VALIDEZ INDIVIDUALES DE LOS SUJETOS: CASO 2 (92 SUJETOS Y 10 ITEMS)

Los 92 sujetos restantes no presentan desajustes. Los valores de validez individuales son aceptables por encontrarse dentro de los límites establecidos.

D.2) FIABILIDAD Y VALIDEZ INDIVIDUALES DE LOS ITEMS: CASO 2 (92 SUJETOS Y 10 ITEMS)

En la Ilustración 5 aparecen los ítems ordenados de mayor a menor desajuste. El ítem con un mayor desajuste es el ítem 9. Pese a que el valor de la MNSQ está dentro de los límites, las ZSTD están por encima del máximo de dos. Sin embargo, al igual que en un caso anterior, siguiendo las indicaciones del manual de Linacre (2012b), puesto que el resto de datos son correctos, puede ignorarse esa desviación y dar por válidas las medidas de ítems.

Ilustración 5. Ítems ordenados según desajuste que presentan (Caso 2)

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S. E.	INFINIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PT-MEASURE CORR.	EXP.	EXACT OBS%	MATCH EXP%	ITEM	
9	404	109	-.79	.09	1.78	4.7	1.82	4.5	A	.39	.57	27.5	35.1	MC3.9
6	183	109	1.02	.11	1.08	.6	1.73	2.8	B	.32	.45	46.8	50.0	MC3.6
5	437	109	-1.11	.10	1.20	1.3	1.21	1.2	C	.51	.54	41.3	41.4	MC3.5
1	368	109	-.49	.09	1.04	.4	1.09	.7	D	.56	.58	32.1	31.8	MC3.1
8	182	109	1.03	.11	.89	-.6	1.03	.2	E	.45	.45	52.3	50.1	MC3.8
2	235	109	.51	.09	.99	.0	1.01	.2	e	.50	.53	46.8	36.1	MC3.2
7	260	109	.31	.09	.95	-.3	.98	-.1	d	.56	.55	33.0	33.2	MC3.7
10	350	107	-.40	.09	.81	-1.6	.78	-1.7	c	.68	.58	36.4	31.5	MC3.10
3	296	109	.04	.09	.76	-2.2	.75	-2.0	b	.65	.58	36.7	31.4	MC3.3
4	318	109	-.12	.09	.66	-3.3	.69	-2.6	a	.76	.58	31.2	32.3	MC3.4
MEAN	303.3	108.8	.00	.09	1.02	-.1	1.11	.3				38.4	37.3	
S. D.	83.9	.6	.69	.01	.30	2.1	.37	2.1				7.7	7.0	

Fuente: Elaboración Propia

E.2) ANÁLISIS DE CATEGORÍAS DE RESPUESTA: CASO 2 (92 SUJETOS, 10 ITEMS)

El análisis de categorías de respuesta lo podemos realizar a partir de los resultados recogidos en la Tabla 10.

Tabla 10. Análisis de categorías (Caso 2)

Categoría		Observado		Media observada	Media esperada	INFIT	OUT-FIT	Umbral de AN-DRICH	Medida de las categorías	
Nom-bre	Mar-ca	Frecuen-cia	%			MNSQ	MNSQ			
1	1	326	30	-1,07	-1,08	1,04	1,18	NONE	(-1,83)	1
2	2	163	15	-0,55	-0,58	0,85	0,73	-0,13	-0,75	2
3	3	199	18	-0,15	-0,13	0,95	1,16	-0,55	-0,07	3
4	4	216	20	0,30	0,30	0,97	0,92	0,01	0,68	4
5	5	184	17	0,72	0,71	1,04	1,33	0,67	(2,06)	5

Fuente: Elaboración Propia

Los datos muestran que los valores obtenidos para las medias observadas son correctos, no sólo porque siguen una estructura creciente sino porque también son próximos a los esperados. Los umbrales de An-drich, sin embargo, aparecen desordenados cuando lo normal es que sean crecientes. La explicación que damos a este desorden es que en la categoría 1 de respuesta existe un número muy elevado de observa-

ciones, alcanzando el 30% del total y provocando que la distribución entre categorías no sea equilibrada. La causa de este desequilibrio puede deberse a que las empresas únicamente hayan seleccionado dos o tres motivos, sin valorar el resto. Esto mismo puede ser la causa de que la correlación para los sujetos sea baja, puesto que los datos en algunos casos no están completos.

Lo ideal en una situación como esta sería volver a enviar el cuestionario con un menor número de categorías o proceder a la recodificación de las ya existentes (por ejemplo unificando las categorías de respuesta 1 y 2). Sin embargo, esto implicaría reenviar la encuesta para comprobar que la nueva codificación es más adecuada, algo inviable actualmente. No obstante, se tendrá en cuenta de cara a futuras investigaciones.

5. CONCLUSIONES

El presente trabajo tiene como objetivo presentar un caso de aplicación de la Teoría de la Medición de Rasch para la validación de constructos. Para ello, en primer lugar, se describe la problemática de medición existente en las ciencias sociales. Seguidamente se describe la utilidad y las aplicaciones de la Teoría de la Medición de Rasch como posible solución a la problemática de medición planteada. Después, a modo de guía, se expone un ejemplo de validación del constructo “motivación para implantar la mejora continua” haciendo referencia a todos los pasos que se han seguido.

A partir del ejemplo se exponen las distintas comprobaciones a realizar para comprobar la validez del constructo. Estas son: análisis de la unidimensionalidad del constructo; análisis de la fiabilidad y validez globales; análisis de la fiabilidad y validez individuales de los sujetos; análisis de la fiabilidad y validez individuales de los ítems; y análisis de las categorías de respuesta.

A partir de las mismas, se va comprobado la calidad global del constructo y se van proponiendo cambios para aumentar la fiabilidad y validez de las medidas. En algunas ocasiones, los cambios son factibles como por ejemplo cuando se propone extraer 17 sujetos de la muestra; otras veces, sin embargo, los cambios no pueden llevarse a cabo. Éste es el caso de la recodificación de categorías de respuesta. De cara a futuras investigaciones sería adecuado la realización de un pretest que permitiera realizar estas comprobaciones e implementar las mejoras que se detecten sin que ello suponga reenviar de nuevo el cuestionario a todas las empresas.

En el ejemplo expuesto, una vez realizadas todas las comprobaciones, se concluye que el constructo “motivación para implantar la mejora continua” es válido. Por lo tanto, la Teoría de la Medición de Rasch se presenta como una metodología rigurosa y minuciosa, útil para proceder a la validación de constructos.

A partir de aquí, dado que el constructo ha sido validado, este podría utilizarse como instrumento de medición generalizado en otras regiones o incluso en otros países, pudiendo realizar comparaciones estandarizadas.

Además, aunque no es objeto de este estudio, una vez validado el constructo podrían realizarse otro tipo de análisis complementarios como son la jerarquización de motivaciones -qué motivos son más importantes y cuáles menos-, la jerarquización de sujetos -qué empresas estuvieron más motivadas-, o el análisis de comportamientos diferenciales -por ejemplo, ¿las empresas más grandes se ven afectadas por algún motivo en mayor medida que las pequeñas o viceversa? ¿Todos los sectores empresariales se mueven por

las mismas motivaciones?, etc. En lo referente a las jerarquizaciones consideramos que la Teoría de la Medición de Rasch va un paso más allá que otras metodologías ya que no sólo analiza la pertenencia de un ítem a un constructo sino que es capaz de valorar la importancia del mismo para los sujetos. Otra ventaja es la medición conjunta de ítems y sujetos, permitiendo el análisis individualizado del comportamiento de estos últimos. Así, una de las aplicaciones más interesantes de la Teoría de la Medición de Rasch es el análisis comparativo individualizado de cada sujeto respecto al conjunto de sujetos. Este análisis, que hasta donde los autores conocen sólo se puede realizar con esta metodología, permite conocer en qué ítems el sujeto presenta un resultado inferior que el esperado en base a lo obtenido por la media del conjunto de sujetos, presentando por lo tanto una desventaja o una debilidad; cuándo presenta un resultado superior al esperado en base a lo obtenido por la media del conjunto de sujetos, obteniendo por lo tanto una ventaja o fortaleza; y cuándo presenta un comportamiento normal o esperado por ser su resultado similar al de la media del conjunto de sujetos. El lector podría ver cierta similitud entre este análisis y el popularmente extendido análisis DAFO.

La validación del constructo como se comentó en el apartado introductorio es sólo el primer paso de muchos otros, de ahí su importancia para la obtención de resultados válidos y fiables. Se espera que el presente trabajo sirva de guía para todos aquellos interesados en la aplicación de la Teoría de la Medición de Rasch.

References

- Alagumalai, S., Curtis, D.D., & Hnugi, N. (2005). *Applied Rasch Measurement: A Book of Exemplars*. The Netherlands: Springer.
- Armistead, C., & Machin, S. (1998). Business process management: implications for productivity in multi-stage service networks. *International Journal of Service Industry Management*, 9(4), 323.
- Armistead, C., & Pritchard, J.P. (1999). Business process management - lessons from European business. *Business Process Management Journal*, 5(1), 10.
- Bawden, R., & Zuber-Skerritt, O. (2002). The concept of process management. *The Learning Organization*, 9(3), 132-138.
- Biazzo, S., & Bernardi, G. (2003). Process management practices and quality systems standards: risks and opportunities of the ISO 9001 certification. *Business Process Management Journal*, 9(2), 149-169.
- Bond, F.G., & Fox, C.M. (2007). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. 2nd edn. New Jersey: Lawrence Erlbaum Associates, Inc.
- Bullita, C. (2006). Geschäftsprozessmanagement bei Siemens Medical Solutions. In *Geschäftsprozessmanagement in der Praxis*. Eds. A. Geschäftsgebiet, H.J. Schmelzer & W. Sesselmann, Hanser, Munich, pp. 475-489.
- Classe, D., & Mundle, J. (1997). Geschäftsprozessgestaltung im Rahmen des CIP-Prozesses bei Bosch. In *Beschleunigung von Geschäftsprozessen*, ed. H.C. Riekhof, Schäffer-Poeschel, Stuttgart, pp. 211-224.

- Clemmer, J. (1994). Process re-engineering and process improvement - not an either/or choice. *CMA Magazine*, 36-39.
- Corallo, A., Margherita, A., Scalvenzi, M., & Storelli, D. (2010). Building a Process-Based Organization: The Design Roadmap at Superjet International. *Knowledge and Process Management*, 17(2), 49-61.
- Davenport, T.H., & Short, J.E. (1990). The New Industrial Engineering: Information Technology and Business Process Redesign. *Sloan management review*, 31(4), 11.
- Elzinga, D.J., Horak, T., Chung-Yee, L., & Bruner, C. (1995). Business process management: survey and methodology. *Engineering Management, IEEE Transactions*, 42(2), 119-128.
- Febles Acosta, J. (Coord) (2008). *Los Modelos de Rasch en Administración de Empresas. Aplicaciones avanzadas*. Santa Cruz de Tenerife: Fundación FYDE-CajaCanarias edn.
- Feltes, P., & Karuppan, C. (1995). Reengineering: getting down to the business of doing business. *Industrial Management*, 37(4), 3-12.
- Furey, T.R. (1993). A six-step guide to process reengineering. *Planning Review*, 21(2), 20.
- González Montesinos, M.J. (2008). *El Análisis de Reactivos con el Modelo Rasch. Manual Técnico A*. Universidad de Sonora, México D.F: Instituto Nacional para la Evaluación de la Educación.
- GropengieBer, F. (1997). Das P/3S-Programm bei Phoenix: Der Weg zur Lean Company. In *Beschleunigung von Geschäftsprozessen*, ed. H.C. Riekhof, Schäffer-Poeschel, Stuttgart, 197-210.
- Hill, F.M. & Collins, L.K. (1998). The positioning of BPR and TQM in long-term organisational change strategies. *TQM Journal*, 10(6), 438.
- Hung, R.Y.Y. (2006). Business Process Management as competitive advantage: a review and empirical study. *Total Quality Management & Business Excellence*, 17(1), 21-40.
- Jones, R. (1994). Improving your key business processes. *The TQM Magazine*, 6(2), 25-29.
- Kohlbacher, M. (2010). The effects of process orientation: a literature review. *Business Process Management Journal*, 16(1), 135.
- Küng, P., & Hagen, C. (2007). The fruits of Business Process Management: an experience report from a Swiss bank. *Business Process Management Journal*, 13(4), 477.
- Lee, R.G. & Dale, B.G. (1998). Business process management: a review and evaluation. *Business Process Management Journal*, 4(3), 214.
- Linacre, J.M. (2002). What do Infit and Outfit, Mean-Square and Standardized mean?. *Rasch Measurement Transactions*, 16(2), 878.
- Linacre, J. M. (2012a). Winsteps® Rasch measurement computer program User's Guide. Beaverton, Oregon: Winsteps.com
- Linacre, J. M. (2012b). Winsteps® Rasch measurement computer program. Beaverton, Oregon: Winsteps.com
- Lindsay, A., Downs, D., & Lunn, K. (2003). Business processes—attempts to find a definition. *Information and Software Technology*, 45(15), 1015-1019.

- Llewellyn, N., & Armistead, C. (2000). Business Process Management: exploring social capital within processes. *International Journal of Service Industry Management*, 11(3), 225-243.
- Lock Lee, L. (2005). Balancing business process with business practice for organizational advantage. *Journal of Knowledge Management*, 5(1), 29-41.
- Marín-García, J.A., & García-Sabater, J.J. (2010). Traducción al castellano de un cuestionario para identificar las conductas de la mejora continua y etapas en el modelo de evolución. *Working papers on Operations Management*, 1(1), 18-26.
- McAdam, R., & McCormack, D. (2001). Integrating business processes for global alignment and supply chain management. *Business Process Management Journal*, 7(2), 113-130.
- Melan, E.H. (1989). Process Management: A Unifying Framework For Improvement. *National Productivity Review*, 8(4), 395.
- Mittermaier, G., & Braun, M. (2004). Geschäftsprozessmanagement bei Infineon. in *Geschäftsprozessmanagement inside*, eds. H. Ellringmann & H.J. Schmelzer, Hanser, Munich, 1-40.
- Ongaro, E. (2004). Process management in the public sector: The experience of one-stop shops in Italy. *The International Journal of Public Sector Management*, 17(1), 81.
- Palmberg, K. (2010). Experiences of implementing process management: a multiple-case study. *Business Process Management Journal*, 16(1), 93.
- Paper, D. & Chang, R. (2005). The state of business process reengineering: a search for success factors. *Total Quality Management & Business Excellence*, 16(1), 121-133.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
- Sandhu, M., & Gunasekaran, A. (2004). Business Process Development in project-based industry. *Business Process Management Journal*, 10(6), 673-693.
- Silvestro, R., & Westley, C. (2002). Challenging the paradigm of the process enterprise: a case-study analysis of BPR implementation. *Omega*, 30(3), 215-225.
- Oreja, J.R. (2005). *Introducción a la medición objetiva en Economía, Administración y Dirección de Empresas: El Modelo de Rasch*. Tenerife: Instituto Universitario de la Empresa de la Universidad de La Laguna.
- Oreja-Rodriguez, J.R. (2008). *La paradoja de Rasch: Medidas y errores*. Santa Cruz de Tenerife: Instituto Universitario de la Empresa (IUDE).
- Pauls, G. (2006). Geschäftsprozessmanagement in einem mittelständischen Unternehmen am Beispiel SCANSONIC. In *Geschäftsprozessmanagement in der Praxis*, eds. H.J. Schmelzer & W. Sesselmann. Munich: Hanser, 492-510.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
- Sanchez, L., & Blanco, B. (2012). El Modelo de Rasch en Dirección de Operaciones. *Working Papers on Operations Management*, 3(2), 35-47.

- Stevens, S.S. (1959). Measurement, Psychophysics and Utility. In *Measurement: Definitions and Theories*, eds. C.W. Churchman & P. Ratoosh. New York: John Wiley.
- Von Davier, M., & Carstensen, C.H. (2007). *Multivariate and Mixture Distribution Rasch Models. Extensions and Applications*. New York: Springer Science + Business Media, LLC.
- Wahlich, S.M. (2004). Prozessorientierte Organisation bei Vaillant Hepworth. In *Geschäftsprozessmanagement inside*, eds. H. Ellringmann & H.J. Schmelzer. Munich, Hanser, 1-40.
- Wright, B.D., & Mok, M. (2000). Rasch models overview. *Journal of applied measurement*, 1(1), 83-106.